Машинное обучение для решения исследовательских и инженерных задач в науках о Земле

Михаил Криницкий

к.т.н., н.с.
Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга климатических изменений (ЛВОАМКИ)

ПЛАН ЛЕКЦИИ

- Обобщенные линейные модели (generalized linear models, GLM)
- Обобщенные аддитивные модели (generalized additive models, GAM)
- Искусственная нейронная сеть

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \eta_i$$

previously on ML4ES

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \eta_i$$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i$$

$$p(\theta_1, x_i) = \text{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

previously on ML4ES

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \eta_i$$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathrm{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \eta_i$$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \text{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная логистическая регрессия

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$$p_k(\theta_1, x_i) = \text{softmax}\big(\eta_i^k, \{\eta_i^k\}\big) = \frac{\exp \eta_i^k}{C^*}$$

$$C^* = \sum_{k=1}^{K} \exp(\eta_i^k)$$

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \eta_i$$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathrm{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная логистическая регрессия

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$$p_k(\theta_1, x_i) = \mathrm{softmax}\big(\eta_i^k, \{\eta_i^k\}\big) = \frac{\exp \eta_i^k}{C^*}$$

Обобщенные линейные модели: модели, в которых некоторая функция $g(\cdot)$ мат.ожидания параметра распределения целевой переменной вычисляется как **линейная функция** признакового описания объектов (событий)

$g(\cdot)$ – т.н. функция связи (link function)

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \eta_i$$

$g = I$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$g = \ln \frac{p}{1-p}$

$$p(\theta_1, x_i) = \mathrm{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная логистическая регрессия

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$g = \ln p + C$

$$p_k(\theta_1, x_i) = \mathrm{softmax}\big(\eta_i^k, \{\eta_i^k\}\big) = \frac{\exp \eta_i^k}{C^*}$$

Обобщенные линейные модели: модели, в которых некоторая функция $g(\cdot)$ мат.ожидания параметра распределения целевой переменной вычисляется как линейная функция признакового описания объектов (событий)

$g(\cdot)$ – т.н. функция связи (link function)

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$g^{-1} = I$

Мультиномиальная логистическая регрессия

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\big(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}\big) = \frac{\exp \eta_i^k}{C^*}$$

$g^{-1}(\eta) = \mathrm{softmax}(\eta_i^k, \{\eta^k\})$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta_i^1}) = \frac{1}{1 + \exp(-\eta_i^1)}$$

$g(\eta) = \sigma(\eta)$

Альтернативно (некорректно, но легче понять):

Обобщенные линейные модели: модели, в которых мат.ожидание параметра распределения целевой переменной вычисляется как некоторая **функция** $\boldsymbol{g^{-1}}(\cdot)$ от линейной функции η признакового описания объектов (событий) x_i

$\boldsymbol{g}(\cdot)$ – т.н. функция связи (link function)

Не изменится

Изменится

Как зависит вид функции потерь от вида функции связи?

$LR:\ y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$

$\mu_i = f(\theta x_i)$

θ

$\mathcal{L} = c\sum(y_i - \mu_i)^2$

Как зависит вид функции потерь от вида функции связи?

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$
$$\eta_i = \theta \cdot x_i$$
$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$\mu_i = g(\eta_i)$

LINE

$$p(y_i, x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right)$$

$$L(\mathcal{T}, \theta) = \prod_{\mathcal{T}} p(y_i, \mu_i)$$

$$\ell(\mathcal{T}, \theta) = \ln L(\mathcal{T}, \theta) = \sum_{\mathcal{T}} \ln p(y_i, \mu_i) = \ln \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\mathcal{T}} \left(-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right)$$

$$\theta = \underset{\Theta}{\operatorname{argmax}}\, \ell(\mathcal{T}, \theta) = \underset{\Theta}{\operatorname{argmin}} \left(\ln \frac{1}{2\sigma^2\sqrt{2\pi\sigma^2}} * \sum_{\mathcal{T}} (y_i - \mu_i)^2\right)$$

$$\mathcal{L}(\mathcal{T}, \theta) = C \sum_{\mathcal{T}} (y_i - \mu_i)^2$$

Как зависит вид функции потерь от вида функции связи?

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta}_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

$$L(\mathcal{T}, \theta) = \prod_{\mathcal{T}} p(y_i, p_i) = \prod_{i=1}^{N} \left(p_i^{y_i} * (1 - p_i)^{(1-y_i)}\right)$$

$$\ell(\mathcal{T}, \theta) = \ln(L(\mathcal{T}, \theta)) = \sum_i \log\left(p_i^{y_i}\right) + \sum_i \log\left((1 - p_i)^{(1-y_i)}\right) =$$

$$= \sum_{\mathcal{T}} (y_i * \log p_i + (1 - y_i) * \log(1 - p_i))$$

$$\theta = \operatorname*{argmax}_{\Theta} \ell(\mathcal{T}, \theta) = \operatorname*{argmin}_{\Theta} \left(-\sum_{\mathcal{T}} (y_i * \log p_i + (1 - y_i) * \log(1 - p_i))\right)$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} (y_i * \log p_i + (1 - y_i) * \log(1 - p_i))$$

$g'(\eta) = \sigma(\eta_i^1)$

$g'(\eta) = \Gamma(\eta_i^1)$

$g'(\eta) = 1 + \frac{1}{2}\tanh(\eta)$

$p_i \in [0; 1]$

Как зависит вид функции потерь от вида функции связи?

НИКАК

вид функции потерь $\mathcal{L}(\mathcal{T},\theta)$ зависит от вида распределения целевой переменной y_i, но не от вида функции связи $g(\cdot)$

θx_i

- обратная функция связи $g^{-1}(\cdot)$ должна отображать R^1 (множество значений произвольной линейной функции $\theta \cdot x_i$) на множество параметров распределения переменной y_i

  $\mu_i \in R^1$ $\quad p_i \in [0,1] \cap R^1$

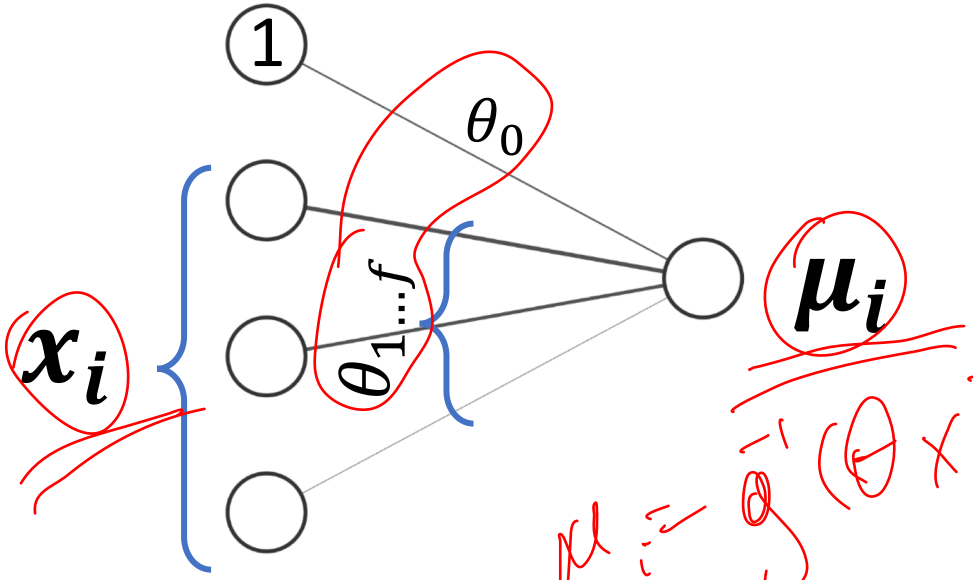
- вычисление параметра (параметров) распределения переменной y_i в модели производится согласно принципу GLM

  $p_i = g^{-1}(\theta x_i)$

- при таких условиях вычисление правдоподобия выборки $\mathcal{T}$ производится независимо от вида функции связи $g(\cdot)$ => вычисление функции потерь в подходе максимизации правдоподобия также производится независимо от вида функции связи $g(\cdot)$

Диаграммы GLM

Линейная регрессия



$\mu = g^{-1}(\theta x_i)$

$$y \sim \mathcal{N}(\mu, \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$$\mathcal{L}(\mathcal{T}, \theta) = C \sum_{\mathcal{T}} (y_i - \mu_i)^2$$

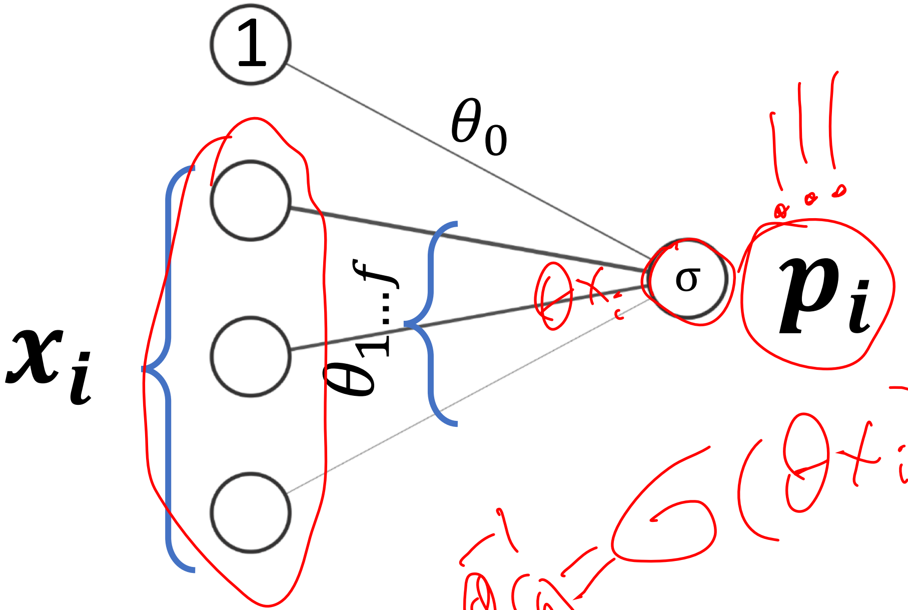
$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = -2X^TY + 2X^TX\theta$$

=> градиентная оптимизация

Диаграммы GLM

Логистическая регрессия



$\hat{y} = \sigma(\theta x_i)$

$$\boldsymbol{y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)}$$

$$\eta_i^1 = \theta_1 \cdot x_i$$

$$p(\theta_1, x_i) = \boldsymbol{\sigma\big(\eta_i^1\big)}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \big(y_i * \log p_i + (1 - y_i) * \log(1 - p_i)\big)$$

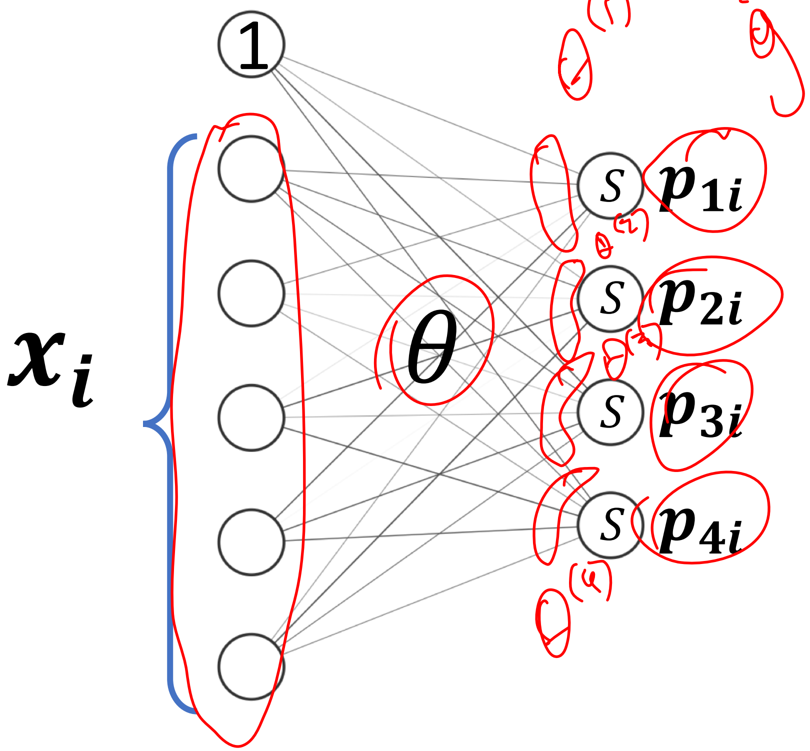
$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = -\sum_i \big(y_i - p(\theta, x_i)\big) x_i$$

=> градиентная оптимизация

Диаграммы GLM

Мультиномиальная логистическая регрессия



$g^{-1}(\eta_k) = S(\eta_k) = \mathrm{softmax}(\eta_k)$

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\big(\boldsymbol{\eta}_i^k, \{\boldsymbol{\eta}_i^k\}\big) = \frac{\exp \eta_i^k}{C^*}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \sum_{k=1}^{K} \big([y_i == k] * \log p_{ik}\big)$$

$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$[True] = 1$

$$\nabla_{\theta_k} \mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \big([y_i == k] - p_{ik}\big) x_i$$

$[False] = 0$

=> градиентная оптимизация

$\exp(\sin(x))$

Регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \to \begin{bmatrix} 1 \\ f_1(x_1) \\ f_2(x_2) \\ f_3(x_3) \end{bmatrix}$$

$$\mu(\theta, x_i) = \eta_i$$

Бинарная классификация

$$y_i \sim \mathcal{B}(p(\theta, x_i))$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln\frac{p_i}{1 - p_i}$$

$$g = \ln\left(\frac{p}{1-p}\right)$$

$$p(\theta_1, x_i) = \mathrm{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

$f_1, f_2, f_3 \dots f_f$ – могут быть разными

Мультиномиальная классификация

$$y_{ik} \sim \mathcal{B}(p_k(\theta, x))$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln p_{ik} + C$$

$$p(\theta_1, x_i) = \mathrm{softmax}(\eta_i^k, \{\eta_i^k\}) = \frac{\exp \eta_i^k}{C^*}$$

$g = \ln(\cdot)$

Обобщенные аддитивные модели – это обобщенные **линейные** модели, в которых некоторая функция $g(\cdot)$ мат.ожидания параметра распределения целевой переменной вычисляется как **линейная функция** некоторых других гладких функций (часто нелинейных, но необязательно) признакового описания объектов (событий)

Регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$g = g^{-1} = I$

Мультиномиальная классификация

$$y_{ik} \sim \mathcal{B}\left(p_k(\theta, x)\right)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln p_{ik} + C$$

$$p(\theta_1, x_i) = \mathbf{softmax}\left(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}\right) = \frac{\exp \eta_i^k}{C^*}$$

$g^{-1} = Sm$

Бинарная классификация

$$y_i \sim \mathcal{B}\left(p(\theta, x_i)\right)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta_i^1}) = \frac{1}{1 + \exp(-\eta_i^1)}$$

$g^{-1} = \sigma$

Альтернативно (некорректно, но легче понять):

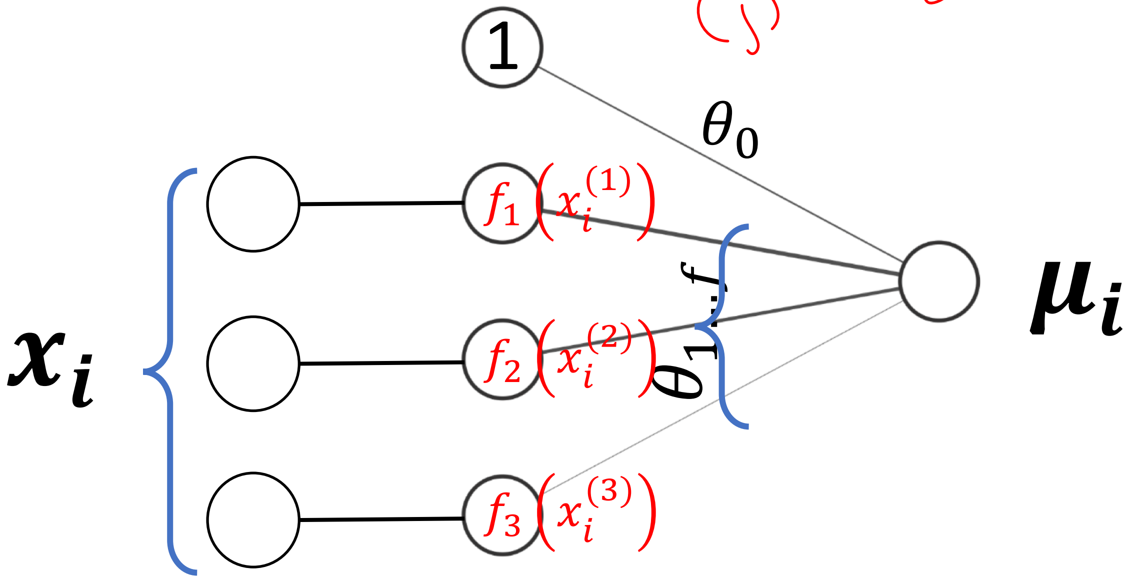
Обобщенные аддитивные модели – это обобщенные линейные модели, в которых в качестве признаков объектов x_i выступают некоторые гладкие (часто нелинейные) функции этих признаков

Диаграммы GAM

Регрессия

i - индекс объекта

(j) - индекс признака



$$\boldsymbol{y \sim \mathcal{N}(\mu, \sigma^2)}$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f(x_i^{(f)})\right]$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$$\mathcal{L}(\mathcal{T}, \theta) = C \sum_{\mathcal{T}} (y_i - \mu_i)^2$$

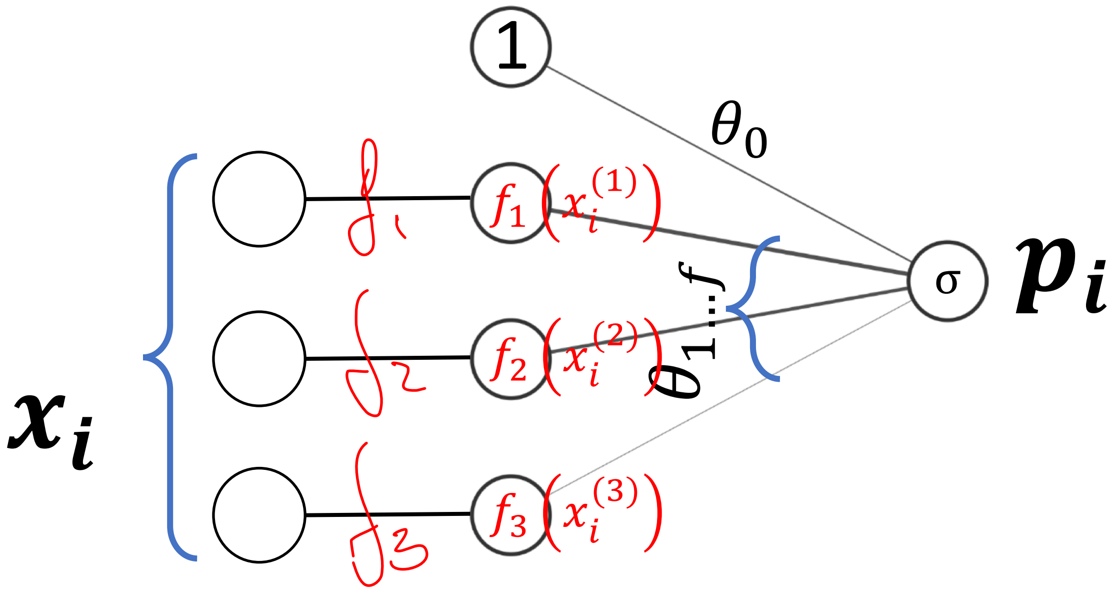
$$\theta = \underset{\Theta}{\operatorname{argmin}}\left(\mathcal{L}(\mathcal{T}, \theta)\right)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = ?$$

=> градиентная оптимизация

Диаграммы GAM

Бинарная классификация



$$\boldsymbol{y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)}$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \ldots f_f\left(x_i^{(f)}\right)\right]$$

$$p(\theta_1, x_i) = \boldsymbol{\sigma(\eta_i^1)}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} (y_i * \log p_i + (1 - y_i) * \log(1 - p_i))$$

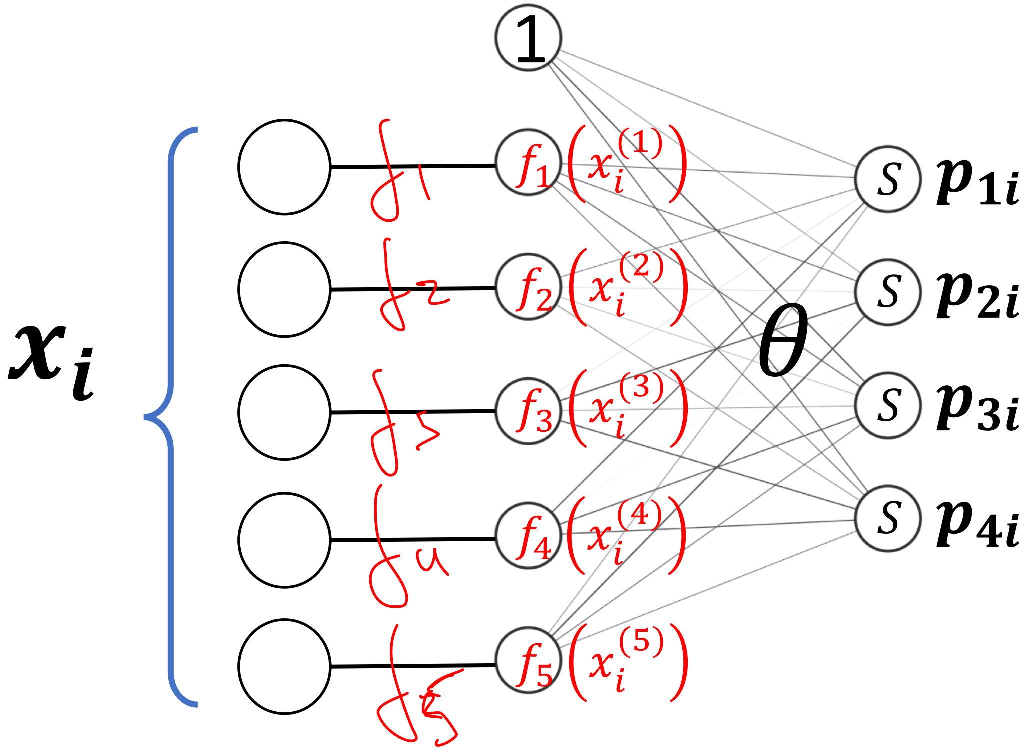
$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = ?$$

=> градиентная оптимизация

Диаграммы GAM

Мультиномиальная классификация



$$y_{ik} \sim \mathcal{B}\left(p_k(\theta, x)\right)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\ln p_{ik} + C = \eta_i^k$$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\left(\boldsymbol{\eta}_i^k, \{\boldsymbol{\eta}_i^k\}\right) = \frac{\exp \eta_i^k}{C^*}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \sum_{k=1}^{K} \left([y_i == k] * \log p_{ik}\right)$$

$$\theta = \underset{\Theta}{\operatorname{argmin}}\left(\mathcal{L}(\mathcal{T}, \theta)\right)$$

$$\nabla_{\theta_k} \mathcal{L}(\mathcal{T}, \theta) = ?$$

=> градиентная оптимизация

Обобщенные аддитивные модели: за и против

ЗА

- Довольно простые, но при этом предоставляют достаточно свободы в выборе нелинейных преобразований $f_j\left(x_i^{(j)}\right)$ исходных признаков объектов (событий);
- Подбирать функции f_j - иногда проще, чем подбирать степени полинома;
- Позволяют применять несколько разных нелинейных функций $f_j\left(x_i^{(j)}\right)$ к разным признакам;
- Модели аддитивные – т.е. можно изучать чувствительность ответа к отдельным входным признакам, просто зафиксировав все остальные.

ПРОТИВ

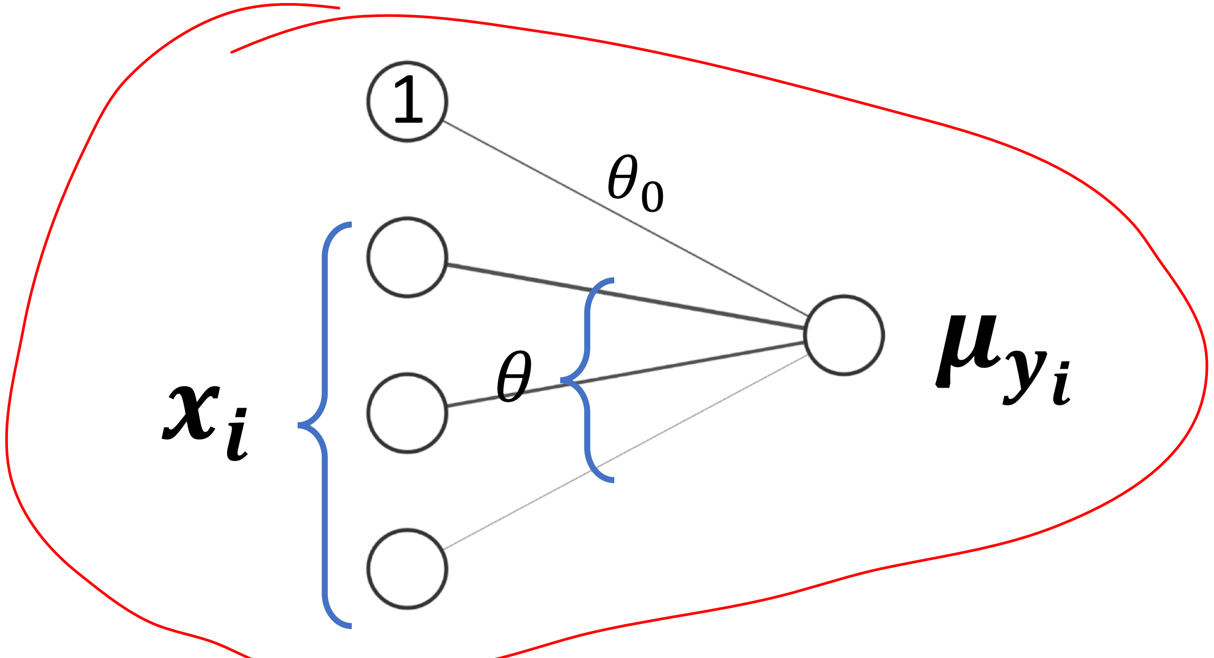
- Это аддитивные модели: учет взаимодействия между признаками ведется только на уровне вычисления линейной функции $g^{-1}(\cdot)$;
- Можно предоставить новые признаки, учитывающие взаимодействие имеющихся, но это не делает сама модель GAM => такие признаки не «обучаемые».

f-ии связи от θx_i

Есть ли способ еще увеличить выразительную способность функциональных параметрических моделей, оставаясь в рамках подхода моделей, обучаемых градиентными методами?

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$ $\mu \in \mathbb{R}^1$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

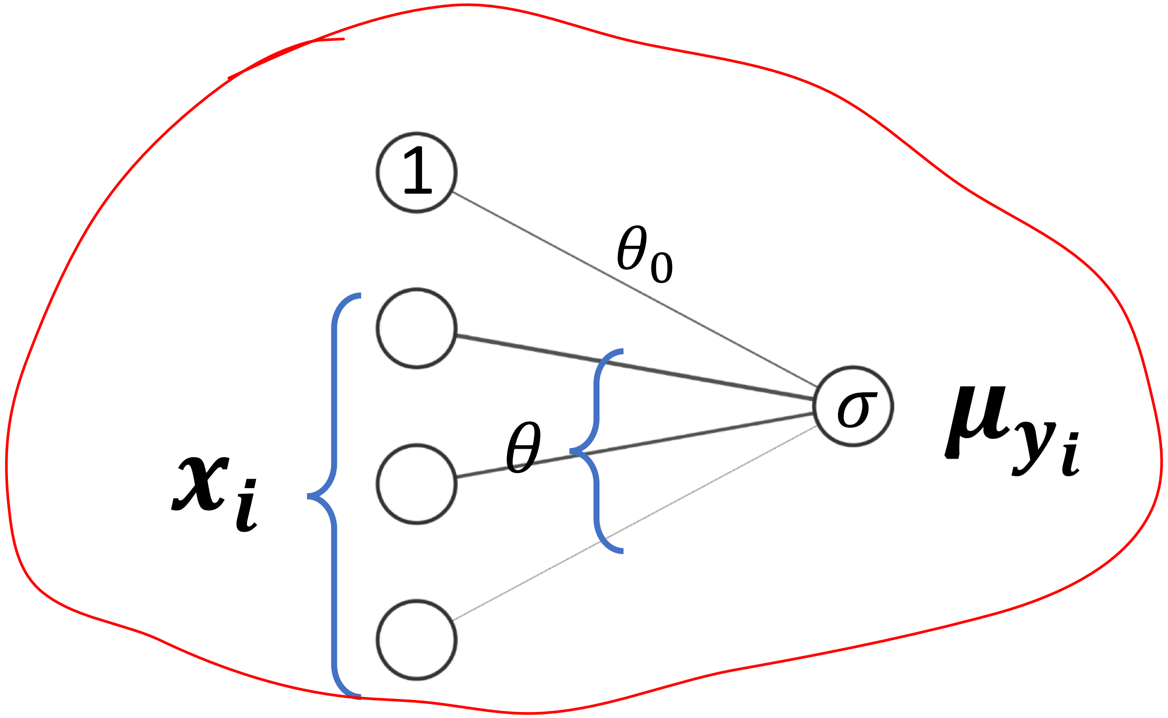


Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

$\sigma = I$ для Л.Р.

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \sigma(\theta \cdot x_i + \theta_0)$

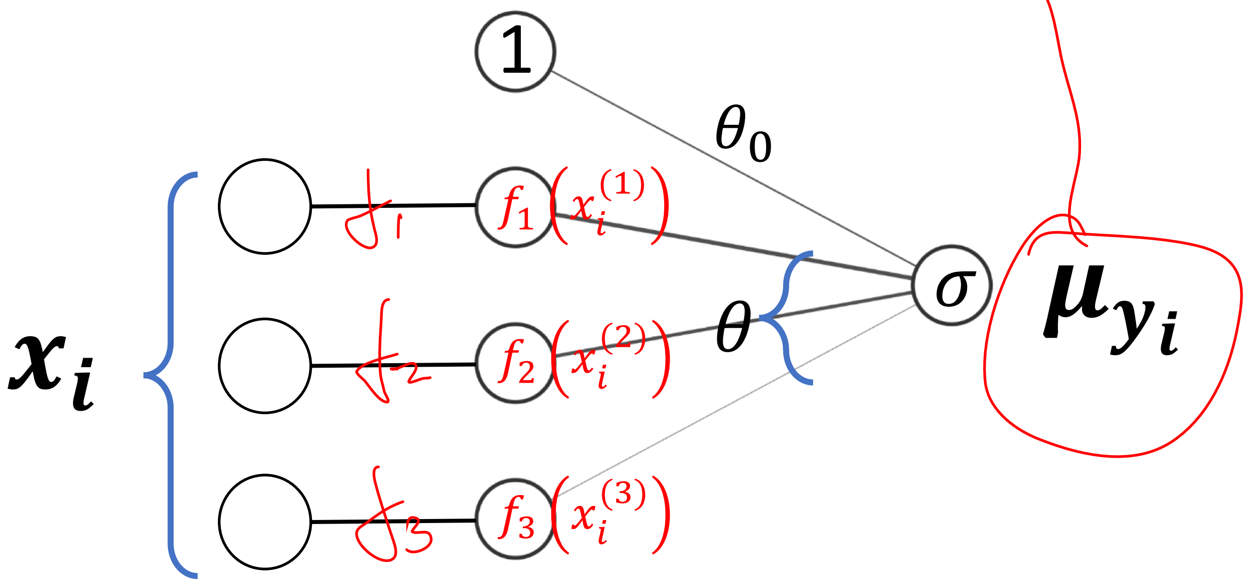


Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \sigma(\theta \cdot x_i + \theta_0)$

GAM: $\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$



Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$

ЛР: $$\mu_{y_i} = \theta \cdot x_i + \theta_0$$

GLM: $$\mu_{y_i} = \phi(\theta \cdot x_i + \theta_0)$$

GAM: $$\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$$

$x_i^{(0)} = 1$

ИНС: $$\mu_{y_i} = \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)$$

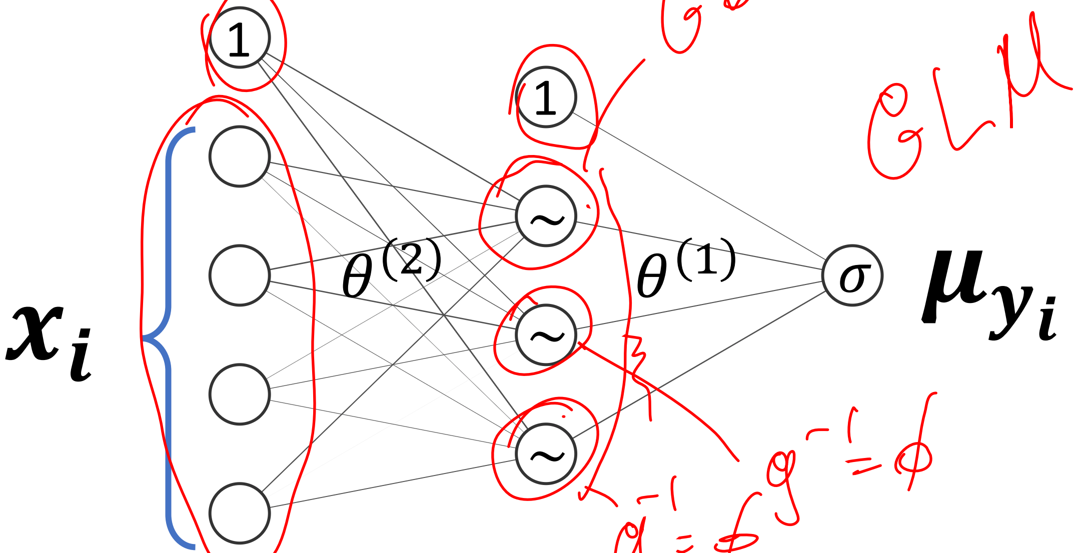
$\phi\left(\theta^{(1)} x_i\right)$

$\xi = \phi\left(\theta^{(1)} x_i\right)$

ϕ: σ, tanh, Γ, sin, cos

GLM $g^{-1} = \phi$

GLM



$g^{-1} = f$ $g^{-1} = \phi$

$\mu_{y_i} = \phi\left(\theta^{(2)} \xi^*\right)$

$\xi^* = \begin{bmatrix} 1 \\ \xi \end{bmatrix}$

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \phi(\theta \cdot x_i + \theta_0)$

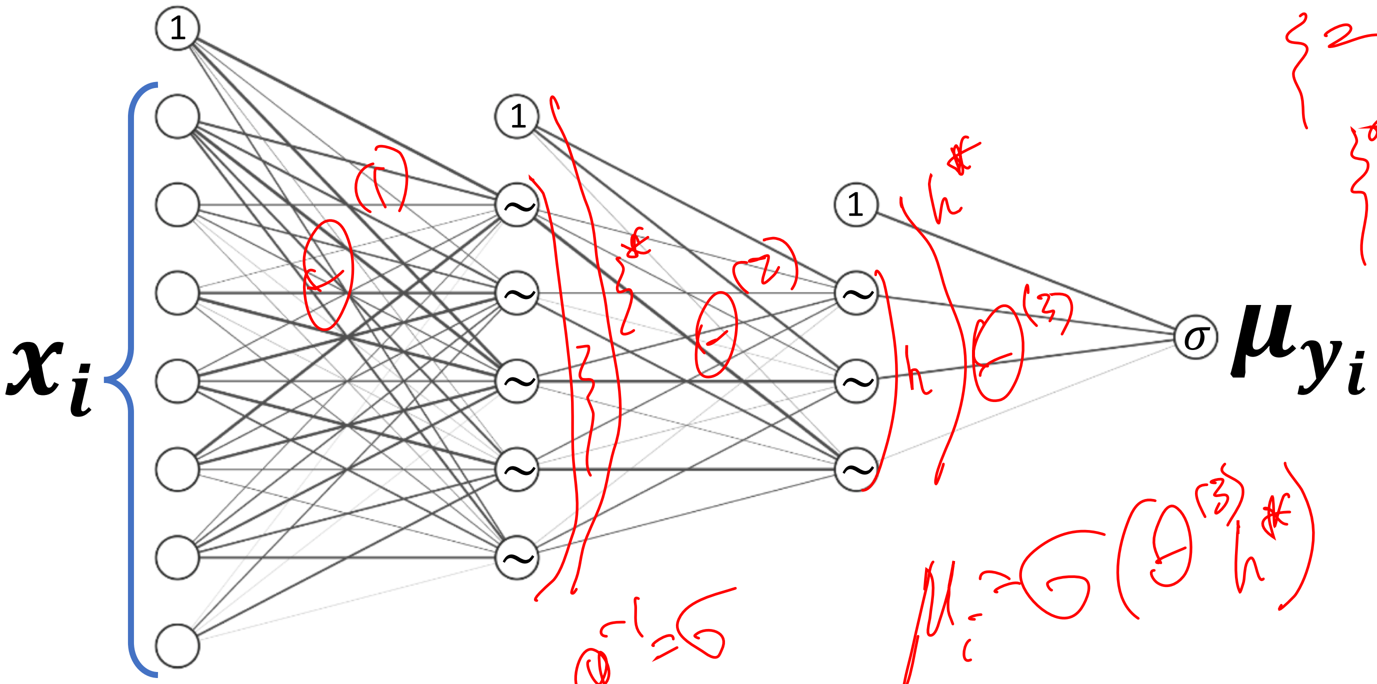
GAM: $\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$

ИНС: $\mu_{y_i} = \phi\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$

ϕ - ф-ция активации

$\xi = \phi^{(1)}(\theta^{(1)} x_i)$

$\xi^* = \begin{bmatrix} 1 \\ \xi \end{bmatrix}$



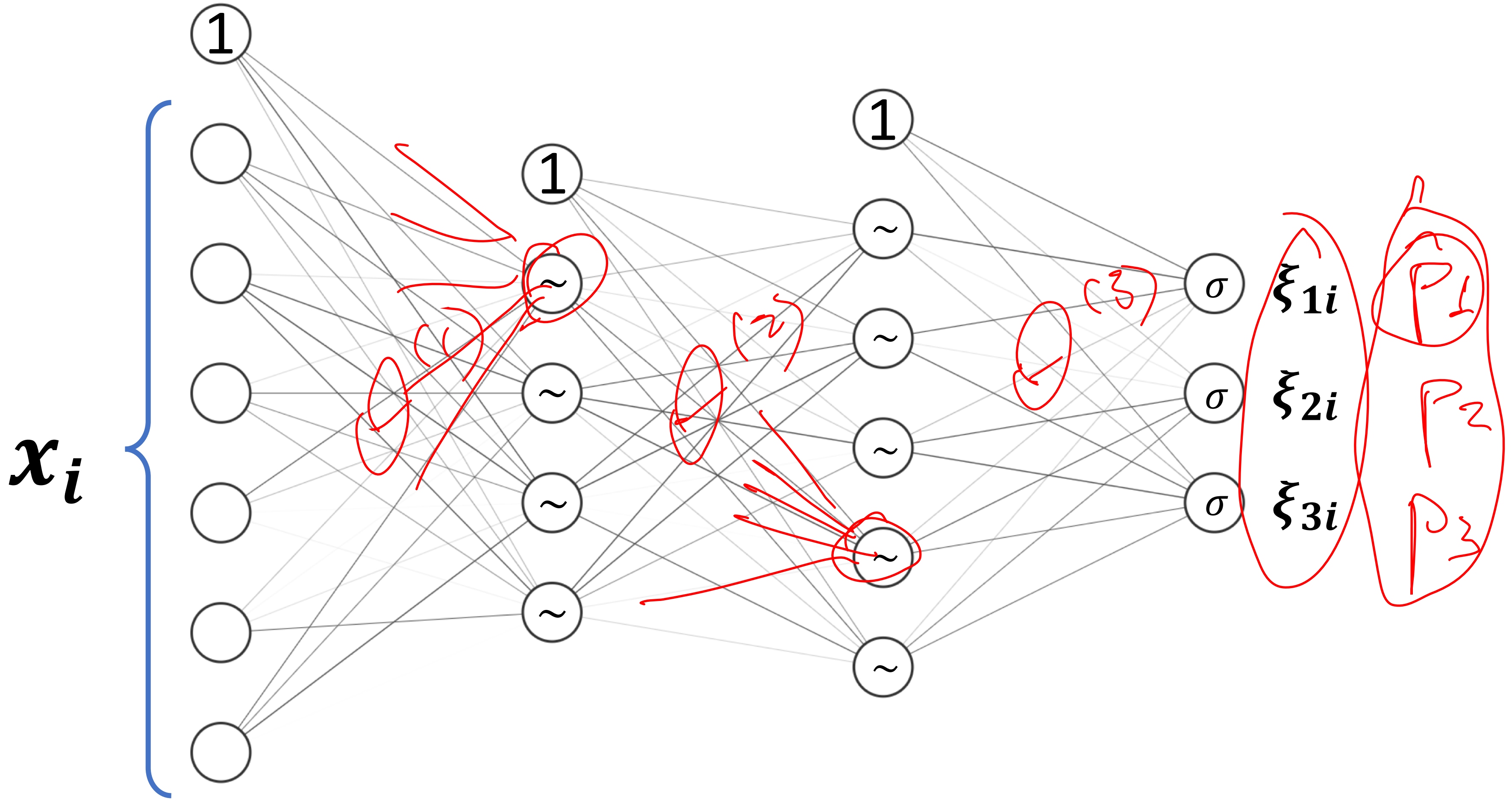
$h = \phi^{(2)}(\theta^{(2)} \xi^*)$

$g^{-1} = \sigma$

$\mu_i = \sigma(\theta^{(3)} h^*)$ $\quad h^* = \begin{bmatrix} 1 \\ h \end{bmatrix}$

Искуственные нейронные сети

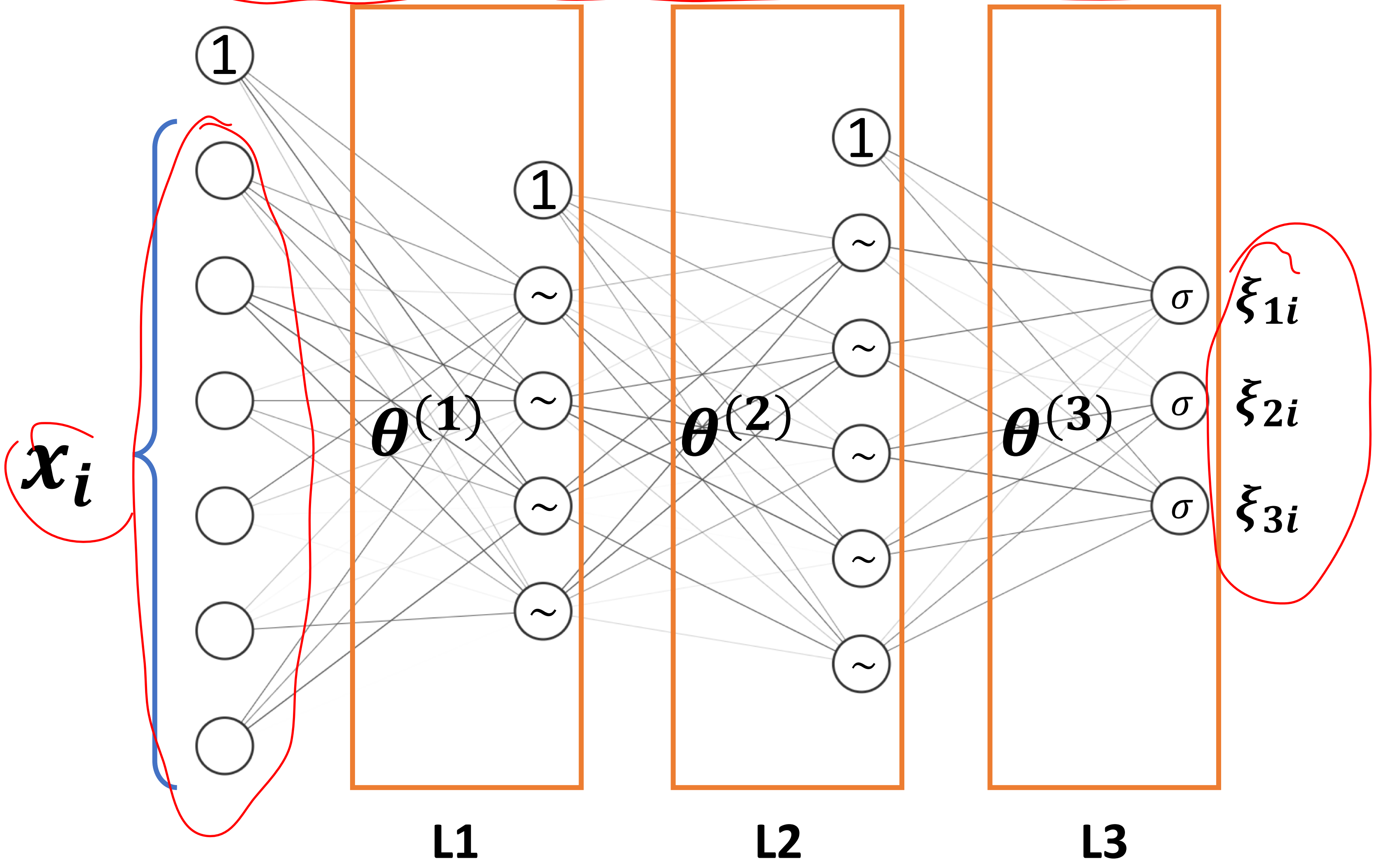
$J = f(f(g(h(x_i))))$



Вид ИНС: многослойный перцептрон (multilayer perceptron, MLP)
Feedforward NN (FNN, сеть прямого распространения)
полносвязная ИНС (Fully-connected NN, FCNN)

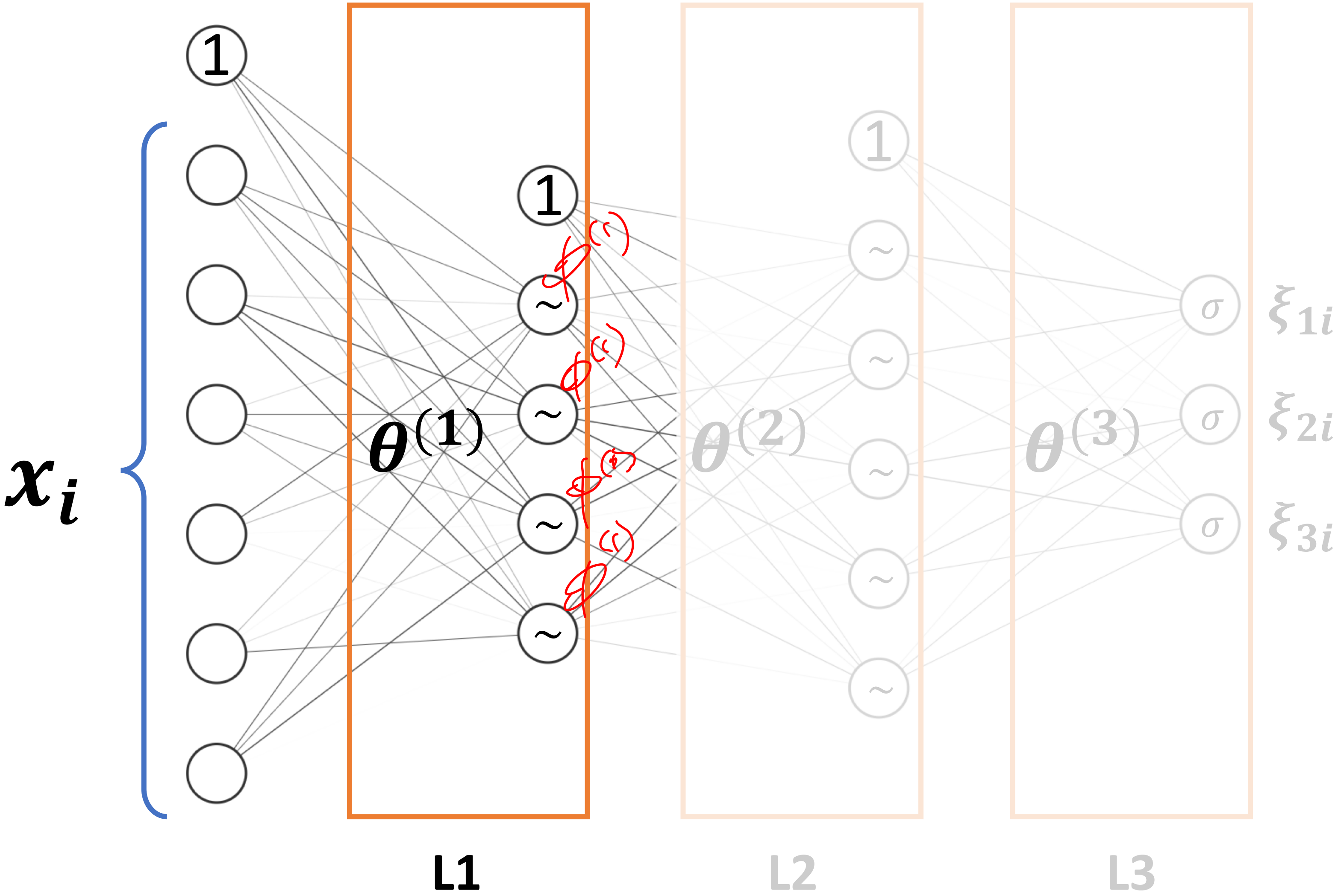
MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$$



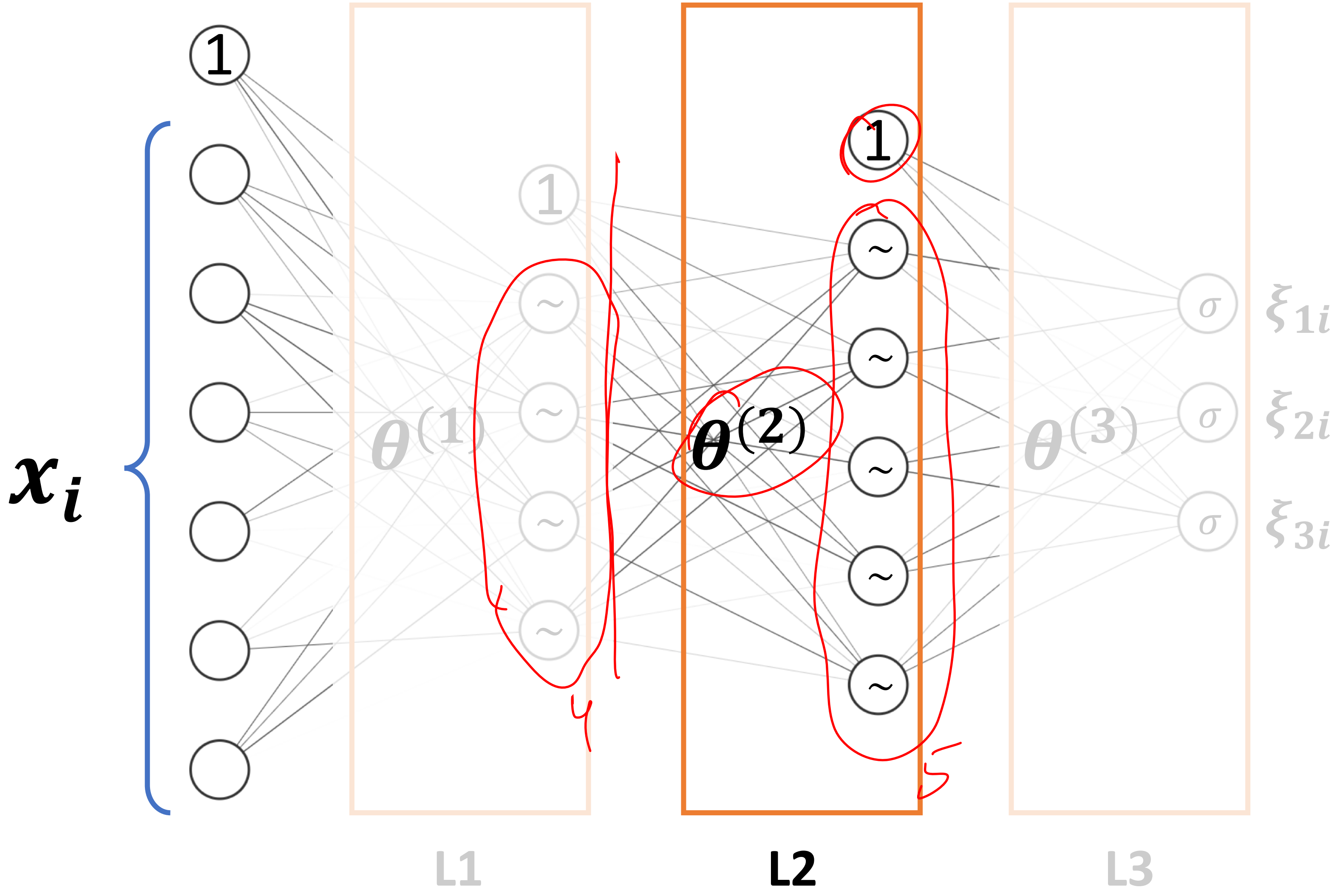
MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi^{(2)}\left(\theta_0^{(2)} + \theta^{(2)} \cdot \boxed{\phi^{(1)}\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)}\right)\right)$$



MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \boxed{\phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)}\right)$$



MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$$